DOCUMENT RESUME

ED 189 110 TM 800 273

AUTHOR Masters, James R.: Dorwart, James P.
TITLE The Development and Validation of a Decision-Making Skills Test.
PUB DATE Apr 80
NOTE 38p.: Paper presented at the Annual Meeting of the National Council on Measurement in Education (Boston, MA, April 8-10, 1980).

EDRS PRICE MF01/PC02 Plus Postage.
DESCRIPTORS Correlation: *Critical Thinking: *Decision Making Skills: Elementary Secondary Education: Grade 5: Grade 8: Grade 11: *Multiple Choice Tests: Predictor Variables: *Problem Solving: *Test Construction: *Test Validity
IDENTIFIERS American School Achievement Tests: Comprehensive Tests of Basic Skills: Cornell Critical Thinking Test: Montana Critical Thinking Test: Pennsylvania: *Pennsylvania Educational Quality Assessment: SRA Assessment Survey

ABSTRACT

Instruments to measure decision-making skills of grade 5, 8, and 11 students were developed as a part of Pennsylvania's statewide assessment program. A decision-making model was formulated based upon John Dewey's five major steps of the critical thinking process: (1) identify the problem: (2) establish the facts: (3) formulate the hypothesis: (4) test the hypothesis: (5) evaluate the results. The model was viewed primarily as an item-writing guide. The general pattern in the instruments calls for the students to read a paragraph describing a decision-making situation involving a central character. Following the presentation of the facts is a series of four-option multiple-choice questions to which students respond. These items measure students' abilities to draw proper inferences from the given facts, to determine appropriate additional data needed by the central character, to predict the consequences of decisions and to match the central character's value system with the decision choices. Validation support for the instruments was gathered in a study in four school districts. (GSK)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED189110

U S DEPARTMENT OF HEALTH, EDUCATION & WELFARE
NATIONAL INSTITUTE OF EDUCATION

THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL NATIONAL INSTITUTE OF EDUCATION POSITION OR POLICY

THE DEVELOPMENT AND VALIDATION OF A DECISION-MAKING SKILLS TEST

James R. Masters
James P. Dorwart

Pennsylvania Department of Education

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY
J. Masters

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

Presented at the Annual Meeting of the
National Council on Measurement in Education

Boston, Massachusetts
April 1980

TM 800273

Summary

Instruments to measure decision-making skills of grade 5, 8 and 11 students were developed as a part of Pennsylvania's statewide assessment program. In these measures the facts of a decison-making situation of a central character are first presented. Students must then respond to multiple-choice items measuring their abilities to draw proper inferences from the given facts, to determine appropriate additional data needed by the central character, to predict the consequences of decisions and to match the central character's value system with the decision choices. Validation support for the instruments was gathered in a study in four school districts.

I. Introduction

In 1965 Pennsylvania adopted ten goals of quality education for its schools. Assessment of the extent to which schools of the Commonwealth were meeting these goals was begun in 1970 by a bureau of the Department of Education called Educational Quality Assessment (EQA).

A number of the ten goals were considered affective in nature.[1] Among these was the tenth one, Adapting to Change. This goal was stated in the following way: "Quality education should help every child to prepare for a world of rapid change and unforseeable demands in which continuing education throughout adult life should be a normal expectation."

In the beginning EQA assessment and in all those until 1978, affective measures were used to assess this goal area. Students were asked both how they felt about change in their lives and how they would react to various change situations. However, these assessment measures never proved highly useful to school district personnel. Also in an era when accountability and the basics are at the forefront, the EQA program was subjected to criticism because of the number of affective tests included in its batteries.

For these reasons the EQA Director requested that the Department's Division of Research develop measures of decision-making skills to replace the affective measures of goal ten in the batteries. This definition was chosen because it was felt by all concerned that decision-making skills were highly important in adapting to change. The decision-making skills tests were developed by James Masters, James Dorwart and Gregory Shannon and first appeared in the EQA batteries for grades 5, 8 and 11 in 1978.

---

[1] Pennsylvania is in the process of changing its goals. Twelve have now been formulated, including one termed "Analytical Thinking."

II. The Development of the Instruments

In developing the new measures, the research literature was first searched to determine whether a decision-making skills test of the type needed already existed. What was required was a paper and pencil test which caused students to think logically about hypothetical decisions.

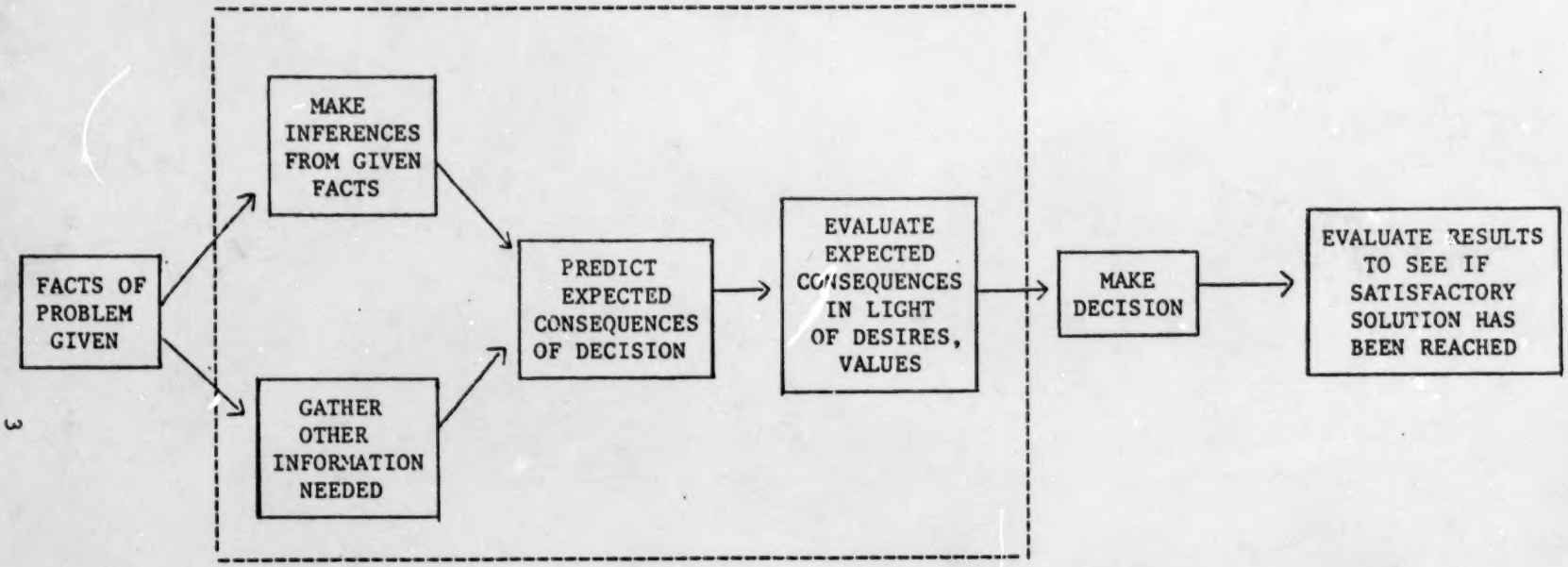
While it was recognized that any verbal skills measure would correlate rather highly with reading comprehension measures, the decision-making measures would be considered of little use if they measured only reading ability. Therefore any test found would not be acceptable if it required either extensive reading or a high level of reading comprehension.

The measures which came closest to what was desired were assessments of critical thinking ability. However, none of those which were found placed respondents in decision-making situations and most required a high level of reading ability.

Thus, a decision-making model was formulated and items were written to match the steps of this model. The model was based upon John Dewey's (1933) formulation of the five major steps of the critical thinking process — 1) Identify the problem; 2) Establish the facts; 3) Formulate hypotheses; 4) Test hypotheses; and 5) Evaluate the results.

The model which guided the development of the test is shown in Figure 1 on page 3. As can be seen in this figure the model begins with the assumption that a problem situation requiring a decision has been presented to the respondent and that the basic facts of this problem have been given. The decision-maker must add to these facts in two ways, first by inferring other truths and second, by gathering additional information. Then, to arrive at a logically reasonable decision, the decision-maker must be able to both predict the consequences of various decisions and evaluate these expected consequences in light of his or her own value system. Taking everything into consideration, a decision is made. The effects of this decision are then evaluated.

Figure 1
DECISION-MAKING MODEL


FACTS OF PROBLEM GIVEN
MAKE INFERENCES FROM GIVEN FACTS
GATHER OTHER INFORMATION NEEDED
PREDICT EXPECTED CONSEQUENCES OF DECISION
EVALUATE EXPECTED CONSEQUENCES IN LIGHT OF DESIRES, VALUES
MAKE DECISION
EVALUATE RESULTS TO SEE IF SATISFACTORY SOLUTION HAS BEEN REACHED

It is not assumed in the model that the sequential approach shown is the way important decisions should be made. In essence, decision-making is considered a complex process and when a logical approach is used it is probably most accurate to assume that the various steps shown can occur in various orders and that cycling and recycling through the steps may routinely occur. For example, a person attempting to make an important decision may begin by listing possible decisions and thinking about their consequences. He or she may then be forced to gather more information before some degree of accuracy of prediction takes place.

The model shown, then, should be viewed primarily as an item-writing guide. The four blocks enclosed by the dotted lines were felt to be abilities important to the making of logical decisions. At the same time it was believed possible that paper and pencil measures of these abilities could be constructed. To develop a test measuring a person's ability to make effective decisions in his or her own life was viewed as too difficult a task, if not an impossible one. The tests developed measure abilities important to logical decision-making. Whether or not these abilities are applied effectively by those possessing them was viewed as a question beyond the scope of the test development undertaking.

The general pattern in the instruments calls for students first to read a short paragraph in which a decision-making situation of a central character or characters is described. At grades 8 and 11 this paragraph may be supplemented by a tabular display of information. The facts of the problem situation are given in the paragraph or the display. A series of four-option multiple choice questions follow the presentation of the facts.

The instruments contain items measuring each of the four decision-making abilities enclosed by the dotted lines in Figure 1. Figures 2, 3 and 4 illustrate the type of item used to measure these.

Drawing Inference items cause students to examine closely the information presented. To answer these items correctly they must be able to distinguish valid from invalid inferences. A typical statement of a Drawing Inferences question is: "What are you most sure about?" Students must then choose which of four inferences about the information presented is most valid.

Information Seeking items measure students' abilities to evaluate the information presented to them. They must determine such things as when more information is needed to solve a problem, which information would be most or least useful and where to obtain the most reliable information. To measure this area, questions of the same or a similar format to that shown in question 2 of Figure 2 are used.

Predicting Consequences items and Matching Values to Consequences items lead students into the final steps of the decision-making model. In the Consequences items students must be able to predict the most likely results of a course of action initiated by the central character to solve his or her problem. Examples of such items are the last grade 5 item of Figure 2 and the second grade 8 item of Figure 3.

Finally, in the Values items grade 8 and 11 students must, after having been presented with a system of values and desires of the central character, use this information to determine which decision would be best for him or her. Such an item is number 2 of Figure 4.

At the grade 5 level this type of format was found to be too complex. So, as shown in item 3 of Figure 2, at grade 5 a simplified version is used requiring students to determine what values are held by the central character which caused him or her to make the decision which was made.

Two separate tryouts of items were undertaken in developing the instruments. First, a pool of 32 grade 5 items and 33 grade 8 and 11 items were written. The

Figure 2

Grade 5 Items

Sue heard a man on TV talk about a big race. All of the children in the race were the same age. Some of the children in the race ate Crunchies cereal before it. These children ran faster than the other children in the race. The man on TV also said that a famous baseball player named Sam Slam eats Crunchies every day. Sue's friend says that no cereal does more for your body than any other one. Sue wants to be able to run faster than anyone else in her class.

(Drawing Inferences)

1. What <u>are</u> you most sure about from the story?

   A. Sam Slam would <u>not</u> be as good at baseball if he did <u>not</u> eat Crunchies.
   B. Most people who eat Crunchies can run fast.
   C. You can't be good at sports unless you eat cereal.
   *D. You can't tell how much Crunchies helps people who play sports.

(Information Seeking)

2. If you were trying to decide whether or not to buy Crunchies, what would be the <u>best</u> question to ask the man on TV?

   A. How long was the race?
   B. How old were the children in the race?
   *C. Could the children who did well in the race run fast before they ate Crunchies?
   D. Did the children who ran slowest in the race know that some of the other children had eaten Crunchies?

(Values)

3. Sue wants to find out more about cereals before she decides which one to eat. What does this seem to show about her?

   A. It is <u>not</u> really important to her to be able to run fast.
   *B. She believes that the more you know about something, the better the decision you can make about it.
   C. She believes that the more you practice something, the better you will be at doing it.
   D. She is willing to risk losing something she has had to work for.

Mary is 12 year old. She reads in a health book that most children her age get about 10 hours of sleep each night. Mary now sleeps about 9 hours each night. She reads also that many adults sleep only about 8 hours each night. She learns that one of our country's greatest scientists needed only about 4 hours of sleep each night.

(Consequences)

4. Mary decides that she wants to be a scientist. Which of the following is true?

   A. She won't be able to become a scientist because she sleeps too much.
   B. She will become a scientist because she now only needs 9 hours of sleep each night.
   *C. You don't know whether or not she will become a scientist.
   D. She could become a scientist but she sleeps too much to become a great one.

*Designates correct answer.

Figure 3

Grade 8 Items

Pat has gained weight, and now can't wear most of her clothes. She has the chance to go on a vacation, but now does not have enough clothes to wear on it. It is hard for her to lose weight since she goes out with friends who always seem to be eating something. She hears a salesman say that studies have shown that adults who took XYZ Diet Pills 3 times a day lost 20 pounds in one month. She is thinking of taking these pills to lose weight.

(Drawing Inferences)

1. Which of the following are you most sure about?

   A. Pat used to be very thin.
   B. If Pat takes XYZ Pills she will lose weight.
   *C. Pat eats when she goes out with friends.
   D. People who take XYZ Pills more often than 3 times a day will lose more than 20 pounds in one month.

(Consequences)

2. Pat decides to go on a diet all the time and not go out with her friends until she loses the weight she had gained. What will happen because of this decision?

   A. She won't be able to lose the weight she wants to lose.
   B. Her friends won't like her.
   *C. She may or may not lose the weight she wants to lose.
   D. She won't be able to go on the vacation.

*Designates correct answer.

Figure 4

Grade 11 Items

Below is information a school counselor has gathered about five careers. Study this information, then answer the questions below.

Career	% of People Trained For Career Who Find Jobs in It	Job Duties	Education Needed	Salary Level	Chance For Advancing To Higher Level Jobs
I	25%	Work with people almost all the time.	4 years of college	Low pay	Excellent
II	50%	Work at desk most of time, with people a small amount.	2 years of education beyond high school	Average pay	Fair
III	50%	Work at desk some of time, with people some of time.	High school education	Low pay	Fair
IV	Almost 100%	Work at desk some of time, with people some of time.	4 years of college	Average pay	Good
V	Almost 100%	Work at desk most of time, with people a small amount.	3 years of education beyond high school	High pay	Excellent

(Information Seeking)

1. Sam has always wanted a job in which he spends most of his time working with people. What is the best question for him to ask about the information in the table?

   A. What kind of training would he need for Career II?
   B. How soon could he advance if he works in Career V?
   *C. If he is trained for Career I, what other jobs could he get?
   D. Why do people in Career II receive lower salaries than people in Career V?

(Values)

2. Linda can hardly wait to graduate from high school and go to work. But she is willing to continue her education beyond high school if she can get a job which will pay well and offer a good chance to advance to higher-level jobs. She would like a job in which she works at a desk. What should she do?

   A. Go into Career III since she doesn't want to go to school anymore.
   B. Go into Career I since there is a good chance to get a higher-level job.
   C. Go into Career IV. If she has to go to school to get what she wants, it gives her the best job for her money.
   *D. Go into Career V even though she would have to go back to school for three years.

*Designates correct answer.

grade 8 and 11 items were tried out in a suburban district with approximately 60 students at each level. The grade 5 items were tried out with 100 students in a district drawing students from suburban and rural areas. Approximately 20 percent of the students at each level were either interviewed or asked to write explanations for their answers. The purpose of this was to test the validity of each item, i.e., to determine that those answering an item correctly had arrived at the correct answer through a logical thinking process.

The grade 5 items fared well in this tryout. They appeared to be eliciting logical thinking and the reading level required was not high. On the other hand, many of the grade 8 and 11 items appeared to be invalid. To add difficulty to these, the item writers had in some cases made the situations of the central characters extremely complex. Students did not appear to be answering items incorrectly because of faulty reasoning. They seemed simply to get lost in all the details of some situations, to misread parts of others and to forget the details of others. A major revision seemed to be needed.

This was accomplished through use of tabular displays, such as that shown in Figure 4. Approximately one-half of the items contained in the instruments prepared for the second tryout relied upon tables to impart the complex details of the situations. For example, in a consumerism situation five brands of record players are presented with five pieces of information tabulated about each. Students must draw inferences from these data, must determine the usefulness to the central character of the information presented and must deal with the complexity of matching the value system of the central character to the consequences of choosing each of the five brands.

The second tryout of the instruments took place in an urban district. Approximately 80 students were involved at each level. Again interviews took place, and a sample of 10 percent of the students were asked to write down the reasoning behind their choices. A total of 52 items were tried out at the grade 5 level and 62 at the grade 8 and 11 levels. This time the items appeared to work well at all three levels.

To test this further, however, 15 Department of Education staff members considered to be above average in intelligence, completed the items. A number of items were dropped because Department of Education data and student interview data seemed to indicate that the judgments involved could lead to more than one right answer. This was expected to be a problem since a rather high percentage of items were evaluative in nature, causing respondents to choose the most or least useful piece of information to a central character or the best decision for the central character.

A combination, then, of item analysis procedures using Henrysson's (1963) index of adjusted item-total correlation and data assessing item validity was used to decide upon the final items for the instruments. These contain 32 items at grades 8 and 11 and 33 items at the grade 5 level.

Shown in the Appendix are item data and total test data for all three instruments. Included are statistics computed in trying out the measures as well as similar information on the tests when administered in the validation study to be described below. Also included for each grade level is item information computed on a random sample of approximately 3000 students participating in Pennsylvania's statewide assessment program in 1978.

As can be seen, for the tryout samples the tests all achieved reliabilities greater than .80. Although an attempt was made to include in the tests only items with adjusted item-total correlations of .30 or greater, a few items at each level with values below this were retained. These items did not appear to be exceptionally weak to the Department of Education judges nor did they seem to be invalid when interviewer sessions with students were held. They were viewed, therefore, as possibly measuring aspects of logical thinking somewhat different from those contained in the majority of the items of each test, but were felt to be not different enough to be dropped completely.

After the tryout minor revisions were made in a number of items, including those with the poorest statistics. These changes were suggested by the Department of Education judges and, in some cases, were simply a matter of replacing a distractor which was seldom chosen by students.

The items retained in the instruments represent a variety of decision-making situations common to the everyday lives of students of the grade levels tested. Three of the 12 situations at the grade 5 level are the following: 1) What to do when a math book is lost the night before a math test; 2) Whether to join a club if it means not being able to be friends with a boy unaccepted by the club; and 3) What to do when an after-school job causes school grades to go down.

Twenty-seven of the items comprising the grade 8 test are also a part of the grade 11 test. The tests place central characters in 11 decision-making situations. These situations deal with such areas as politics (which candidate to vote for), consumerism (which record player to buy), personal matters (whether to cut long hair to get a job), civic affairs (where to build a new park), health matters (whether to take diet pills), careers (which one to choose) and school matters (which courses to take).

III. The Validation Study

Design of the Study

The purpose of the validation study was that of determining the extent to which the constructed EQA decision-making skills measures related to established measures of similar abilities. Since the instruments which seemed to match the content of the measures most closely were critical thinking tests, the validation study focused upon the extent to which the new scales correlated with established measures of critical thinking. Also, teachers were asked to rate the critical thinking abilities of their students and these ratings were correlated with the EQA measures. Finally, to determine the relationship between the new measures and reading ability, reading scores were gathered from districts participating and were correlated with EQA scores.

Sample

The study took place in four school districts. Students from one rural district (District 1) participated at grades 5, 8 and 11. A second district (District 2),

drawing students from rural and blue-collar homes, took part at grades 8 and 11. And, two suburban districts (Districts 3 and 4) were involved at the grade 5 level only. A total of 182 grade 5 students, 197 grade 8 students and 207 grade 11 students participated.

Instrumentation

The critical thinking measures employed as validation criteria were the Cornell Critical Thinking Test, Level X, (Ennis and Millman, 1971) and the Montana Critical Thinking Test (both grade 6 and grade 12 versions). Each of the districts participating used a different measure of reading achievement.

The Cornell Critical Thinking Test is made up of 71 items. The test is in the form of a story about the adventures of a group of explorers whose spaceship has landed on a planet called Nicoma. They are to investigate the disappearance of a group which had landed on the planet two years previously.

Multiple-choice format is employed in the Cornell Test. The first section (Section A) measures Induction; students are asked to assess the degree to which information gathered supports a specific conclusion. In Section B, Reliability, respondents must determine which of two sources of information is the more reliable. Part II, Section A, is termed Deduction. In these items respondents must determine which of three conclusions follows logically from the reasoning presented. Finally, Part B of Section II measures Finding Assumptions. Students must decide which of three assumptions must have been made in order for a statement made by a central character to be true.

The Cornell Test was administered to students at all three grade levels of the study. However, the tasks required were too complex for many fifth graders. Thus, only about two-thirds of the grade 5 students participating were able to complete the instrument and in one district it was felt to be so difficult that it had to be read to those taking it. Therefore, because of the problems encountered in using the instrument at the grade 5 level, the Cornell was used as a validation criterion only at the grade 8 and 11 levels.

The Montana scales used are included in Montana's statewide assessment program. Both scales employ a multiple-choice format. The grade 6 measure was used in the study as a validation criterion at all three of the grade levels involved. It consists of 17 items measuring students' abilities to determine whether a given statement is fact or opinion.

The grade 12 Montana measure is made up of 21 items. Two types of questions are included. No names for these are given in the Montana manuals, but for purposes of description they will be termed Drawing Inferences (9 items) and Proving Conclusions (12 items). The Drawing Inferences items are similar to those of the EQA scales. A paragraph is presented and students must draw inferences about the information given to determine whether each of nine statements is true or false. If the statement cannot be inferred as either true or false, students are directed to mark "N" for No Inference. The Proving Conclusions subscale requires students to choose which of three statements would be most difficult to prove true or false. The grade 12 measure was used as a validation criterion at both the grade 8 and the grade 11 levels.

The following standardized tests were used by the districts involved:

District 1--American School Achievement Test (Reading, Part 1, Form X), 1963 version. The Reading score is derived from Word Meaning and Paragraph Meaning subscales. The Intermediate level was given to grade 5 students; the Advanced level was used at the grade 8 level. Raw scores from these testings were used in the analyses of the study.

District 2--SRA Assessment Survey (Reading), 1971 version. The Reading score is derived from the sum of the Comprehension and Vocabulary sections. Grade 8 students took the Multilevel edition; grade 11 students were administered the ITED edition. Only percentile scores were gathered from these testings. In order for statistical analyses to be carried out, these were converted to Normal Curve Equivalents (NCE's) using tables presented in manuals for the ESEA Title I Evaluation and Reporting System.[1]

[1] See, for example, Talmadge, G. Kasten and Christine T. Wood. User's Guide, ESEA Title I Evaluation and Reporting System, Mountain View, CA: RMC Research Corp., 1976.

District 3--CTBS Reading, 1969 version. The Reading score is the sum of the scores of the Vocabulary and Comprehension sections. Level 2, Form Q had been administered previously to the grade 5 students involved. Raw scores were obtained and employed in the analyses of the study.

Teacher ratings of the critical thinking abilities of their students were gathered using two types of rating scales (these are shown in the Appendix). The definition of critical thinking given teachers matched closely the model upon which the EQA decision-making skills tests were based.

To produce ratings for the grade 5 students of District 1 and for all eighth and eleventh graders, teachers were instructed to choose from among those listed five to ten students high in critical thinking ability and five to ten students low in this ability. This approach was used to increase the reliability of ratings since at the grade 8 and 11 levels no one teacher could have been chosen who had had enough contact with the students tested to provide a reliable rating. This same situation existed for the fifth graders of District 1 since a departmentalized approach to grade 5 teaching was used there.

Teachers at Districts 3 and 4 were asked to rate the critical thinking abilities of each of their fifth graders using a five-point scale. A departmentalized approach was not used at these districts, so each teacher involved was well aware of the thinking abilities of the class of students assigned to him or her.

Analysis

The major statistical procedure used to analyze the results of the study was correlation. Significant positive correlations between EQA instrument scores and both scores on the other critical thinking measures and teacher ratings of critical thinking were considered evidences of the validity of the EQA scales.

To analyze the high-low teacher rating results, a correlational procedure statistically equivalent to carrying out a t-test was used. Because rating procedures were not constant across districts (some students could have been rated by

as many as five teachers, others by only one), it was decided for analysis purposes to include in the high-rated group those students who were selected for this group by at least one teacher with no other teacher placing them in the low group. A similar procedure was used to decide which students to include in the low group. Those included in the high group were arbitrarily assigned a rating score of "2," those in the low group received a "1." A significant positive correlation between this rating variable and the EQA instrument indicated, therefore, that students rated high in critical thinking tended to score higher on the EQA measure than students rated low. Since the number of students participating in the study at each grade level was large, the .01 level was employed as the level required for significance of correlations.

It was assumed in the study that significant relationships between EQA instrument scores and criterion scores could have been found in part because all the tests used involved reading, and teacher ratings also could have been influenced by teacher perceptions of student reading abilities. Thus, reading scores were collected as a means of determining the extent to which reading entered into all relationships. Statistically, a procedure termed partial correlation was used to take out the effects of reading ability from all other relationships. If, after partialling out reading, it was found that significant relationships did not exist between the EQA instruments and the criterion measures, it would then have to be concluded that the major component they had in common was reading ability.

Reliability Results

Table 1 shows the total test statistics achieved by the EQA measures, the Montana scales and the Cornell Test of Critical Thinking.

Grade 5 Results

The correlations computed between EQA scores and criterion measures are shown in Table 2. As can be seen, the EQA measure correlated significantly with both the Montana scale and teacher ratings. As expected, scores on the EQA measure and the ceiterion measures correlated significantly with reading scores. The fact that Student Sex did not correlate significantly with EQA scores indicated that the test did not appear to favor either boys or girls.

Table 3 shows relationships between the EQA measure and the criterion measures after reading scores have been partialled out. In general, significant relationships remained.

Table 1

Total Test Statistics for the Instruments Used

Grade	No. of Students Tested	Test	No. of Items	Mean No. Correct	Standard Deviation	Standard Error of Meas.	Reliability[1]	Mean Converted Score[2]
5	179	EQA	33	23.07	5.56	2.39	.82	-
	179	Montana (elem. version)	17	12.02	2.85	1.68	.65	-
8	191	EQA	32	16.72	5.82	2.52	.81	-
	190	Montana (elem. version)	17	13.21	2.80	1.52	.70	-
		Montana (sec. version)	21	10.99	3.75	2.10	.69	-
		Cornell	71	36.24	8.55	3.78	.81	17.24
11	197	EQA	32	20.64	5.54	2.41	.81	-
		Montana (elem. version)	17	14.79	2.30	1.31	.67	-
		Montana (sec. version)	21	23.41	3.62	1.96	.71	-
		Cornell	71	41.20	8.18	3.62	.80	26.30

[1]Kuder-Richardson #20.

[2]Scores on the Cornell test are obtained through use of the following conversion:

Total Score = (# Correct) - ½ (# Incorrect).

Table 2

Relationship Between the Grade 5 EQA Instrument and Criterion Measures

	Teacher Rating				EQA				Montana (Elem.)				Student Sex	
	District			Total	District			Total	District			Total		
	1	3	4	Group	1	3	4	Group	1	3	4	Group	1	3
EQA	.78*	.58*	.62*	.64*	-	-	-	-	-	-	-	-	-	-
Montana (Elementary)	.66*	.34*	.63*	.52*	.64*	.60*	.63*	.64*	-	-	-	-	-	-
Student Sex[1]	.08	.11	-.11	.04	-.07	.09	-.14	-.04	-.09	-.07	-.15	-.09	-.17	-.05
Reading	.80*	.75*	-	-	.59*	.73*	-	-	.43*	.47*	-	-	-	-

[1]Positive significant correlations indicate that girls tended to score higher than boys.

*Significant at the .01 level.

Table 3

Relationship Between the Grade 5 Instrument and Criterion Measures with Reading Achievement Partialled Out

	Teacher Rating		EQA	
	District		District	
	1	3	1	3
EQA	.65*	.06		
Montana (Elem.)	.59*	-.02	.53*	.43*

*Significant at the .01 level.

An exception to this occurred for District 3. Neither the Montana nor the EQA correlated significantly with teacher ratings after reading was partialled out.

The reasons for these differing results is difficult to determine because in District 3 teachers used a different rating scale for rating students. Also, each student in District 3 was rated by only one teacher while in District 1 each student was rated by several teachers. The fact that the correlation between teacher ratings and EQA scores was lower in District 3 ($r = .58$) than in District 1 ($r = .78$) may indicate that District 3 teacher ratings were less reliable than District 1 teacher ratings. Additionally, there was a higher correlation between reading scores and EQA scores in District 3 ($r = .73$) than in District 1 ($r = .59$).

Grade 8 Results

The correlations computed between grade 8 EQA scores and criterion measures are shown in Table 4. Except for the variable Student Sex all correlations were significant by district and for the total group. These results all support the validity of the EQA measure.

As was the case in grade 5, when reading scores were partialled out, significant relationships between the EQA instrument and the criterion measures remained. These results are contained in Table 5.

Table 4

Correlations Between the Grade 8 EQA Instrument and Criterion Measures

	Teacher Ratings			EQA			Cornell			Montana						Student Sex		
										Elementary			Secondary					
	District		Total	District		Total	District		Total	District		Total	District		Total	District		Total
	1	2	Group	1	2	Group	1	2	Group	1	2	Group	1	2	Group	1	2	Group
EQA	.75*	.57*	.66*															
Cornell	.70*	.62*	.65*	.73*	.64*	.68*												
Montana (Elem.)	.59*	.59*	.60*	.62*	.60*	.60*	.60*	.52*	.51*									
Montana (Sec.)	.71*	.52*	.61*	.72*	.56*	.63*	.60*	.60*	.59*	.41*	.51*	.54*						
Student Sex	.28	.45*	.38*	.41*	.40*	.41*	.24*	.27*	.24*	.14	.42*	.35*	.30*	.30*	.35*			
Reading	.75*	.49*	-	.72*	.65*	-	.64*	.63*	-	.50*	.59*	-	.60*	.57*	-	.24*	.13	-

*Significant at the .01 level.

[1]Positive significant correlations indicate that girls tended to score higher than boys.

Table 5

Relationships Between the Grade 8 EQA Instrument and Criterion Measures With Reading Achievement Partialled Out

	Teacher Rating		EQA		Cornell		Montana (Elem.)	
	District		District		District			
	1	2	1	2	1	2	1	2
EQA	.46*	.37*						
Cornell	.44*	.45*	.51*	.40*				
Montana (Elem.)	.38*	.42*	.44*	.36*	.42*	.24		
Montana (Sec.)	.45*	.33	.47*	.30*	.32*	.38*	.41*	.83*

*Significant at the .01 level.

Grade 11 Results

The grade 11 results, summarized by Tables 6 and 7, support the validity of the EQA measure. First of all, in Table 6 all correlations between the EQA measure and the criterion measures are significant.

One somewhat negative result shown in Table 6 was a significant correlation between Student Sex and the EQA instrument for the total group tested. This correlation indicated that girls tended to score higher on the EQA scale than boys. However, the correlation was of low magnitude (.21) and if it is an indication that the test is sex-biased, it is also an indication that it is biased in a very minimal way. Also, the significant correlation (.42) between teacher ratings and Student Sex may be evidence that, for the particular sample tested, girls would be expected to score better than boys on critical thinking tasks.

As can be seen in Table 7, when reading was partialled out, significant relationships remained between the EQA instrument and all criterion measures. These relationships ranged from .39 for the Montana (elem) to .54 for both the Montana (sec) and teacher ratings. With the exception of the correlation between teacher ratings and the Cornell, the criterion measures also retained their significant relationships. The EQA measure and the criterion measures, then, seem to share something which is not solely reading ability. This shared factor is assumed to be logical thinking ability.

IV. Statewide Assessment Results

The measures developed are now included in the EQA assessment batteries. Since a matrix sampling approach is employed it is not possible to obtain further information on the tests' statistics. However, it is possible to investigate correlations of the tests with other measures.

Table 8 below summarizes such results. Since the EQA tests are divided into three forms at the grade 5 level and four forms at grades 8 and 11, each grade 5 student responds to 11 decision-making skills items and each grade 8 and 11

Table 6

Correlations Between the Grade 11 EQA Instrument and Criterion Measures

	Teacher Ratings			EQA			Cornell			Montana Elem.			Montana Sec.			Student Sex		
	District 1	District 2	Total Group	District 1	District 2	Total Group	District 1	District 2	Total Group	District 1	District 2	Total Group	District 1	District 2	Total Group	District 1	District 2	Total Group
EQA	.81*	.66*	.71*															
Cornell	.74*	.33	.53*	.62*	.66*	.65*												
Montana (Elem.)	.58*	.60*	.59*	.60*	.60*	.60*	.54*	.58*	.57*									
Montana (Sec.)	.69*	.51*	.59*	.67*	.69*	.69*	.53*	.60*	.58*	.60*	.67*	.65*						
Student Sex[1]	.46*	.37*	.42*	.24	.22*	.21*	.00	-.02	-.01	-.13	.14	.03	.10	.14	.11			
Reading	-	.45*	-	-	.62*	-	-	.53*	-	-	.57*	-	-	.53*	-	-	.16	-

*Significant at the .01 level.

1 Positive significant correlations indicate that girls tended to score higher than boys.

Table 7

Relationships Between the Grade 11 EQA Instrument and Criterion Measures With Reading Achievement Partialled Out (District 2 Only)

	Teacher Ratings	EQA	Cornell	Montana (Elem.)
EQA	.54*			
Cornell	.11	.50*		
Montana (Elem)	.46*	.39*	.40*	
Montana (Sec)	.36*	.54*	.45*	.53*

*Significant at the .01 level.

student answers eight of such items. A z score distribution of individual scores was developed at each grade level and these z scores were correlated with other measures to form Table 8. These correlations are based upon the scores of approximately 3000 students at each level, randomly chosen from those participating in the statewide assessment program in 1978.

Table 8

Correlations of Decision-Making Skills Tests with Other Measures
(Sample = approximately 3000 at each level)

	Parent Occupation[1]	Parent Education	Student Sex[2]	Degree of Urbanness of Home Area[3]	Student Race[4]	Student Reading Comprehension
Grade 5	.18	.18	.12	.02	-.14	.60
Grade 8	.17	.20	.15	-.01	-.09	.56
Grade 11	.14	.20	.18	-.06	-.15	.56

[1]Highest values are given to occupations requiring the most education and bringing in the highest incomes.

[2]Girls receive values of 2, boys receive 1's.

[3]Highest values are given to most highly populated areas.

[4]White students received the highest values.

As shown in Table 8 the decision-making skills tests did not correlate highly with measures of socioeconomic status, with degree of urbanness of students' home areas or with the sex or race of students. Correlations with reading comprehension were much higher, but low enough that the measures would not be considered to be measuring the same thing as the reading comprehension test. It must be remembered in interpreting all these results, however, that each student answered only a part of the decision-making skills items. Lowered reliability of the tests could therefore have some effect upon the results.

V. Conclusions

The EQA decision-making skills tests were developed to fulfill a need of Pennsylvania's statewide testing program. A model was formulated to guide this development effort.

The instruments which were constructed appear to possess a number of desirable measurement qualities. Their reliabilities are adequate and evidences that they are measures of thinking have been obtained.

A valid criticism of the new measures is that it is difficult to pinpoint exactly what they are measuring in comparison with other similar instruments already developed. Is it decision-making ability? Is it the ability to think logically? Is it critical thinking ability? Is it critical reading? Is it information processing? Or is it some combination of factors?

An examination of the research literature was of little help in clarifying the situation. This is because there is little agreement about definitions of terms. Definitions of critical thinking, for example, vary greatly.

As another clarification attempt it was planned in the validation study to correlate subtests of the criterion measures and the EQA measures. A possible factor analysis of these correlations could then have been undertaken.

However, the reliabilities of these subtests were so low that correlations among them would have been of dubious value. At the grade 11 level, for example, three of the four Cornell subtests had reliabilities below .60.

Much work, therefore, remains to be accomplished, both in the construction of measures such as those described in this paper and in the clarification of definitions of the terms being used. Many educators would view as valid goals of schools the development of logical thinking abilities, of critical thinking abilities or of decision-making skills. To adequately document our attempts to achieve these goals, clear definitions of these terms and widely accepted measures of them must be produced.

References

Dewey, J. How We Think, Boston: D. C. Heath and Company, 1933.

Ennis, R and Millman, J. Manual for the Cornell Critical Thinking Test, Level X and Cornell Critical Thinking Test, Level Z, Urbana, Illinois: University of Illinois, 1971.

Henrysson, S. Correction of Item-Total Correlations in Item Analysis. Psychometrika, 1963, 27, 419-424.

APPENDIX

Total-Test and Item Statistics for the Tryout Sample, the Validation Sample and the Statewide Assessment Sample

Total Test Statistics

	Grade 5		Grade 8		Grade 11	
	Tryout Sample	Validation Sample	Tryout Sample	Validation Sample	Tryout Sample	Validation Sample
Number of Respondents	79	173	75	179	83	188
Total Score Mean	17.82	23.07	16.63	16.72	18.80	20.64
Total Score Standard Deviation	6.16	5.56	6.24	5.82	6.27	5.54
Standard Error of Measurement	2.62	2.39	2.50	2.52	2.45	2.41
Est. Inter-Item r	.12	.12	.14	.12	.15	.12
Reliability (Kuder-Richardson #20)	.82	.82	.84	.81	.85	.81

Summary of Adjusted Item-Total Correlations

	Numbers of Items with Values in Each Range					
	Grade 5		Grade 8		Grade 11	
Range	Tryout Sample	Validation Sample	Tryout Sample	Validation Sample	Tryout Sample	Validation Sample
.25 or below	3	7*	6	6	5	5
.26 - .50	25	23	20	21	22	26
.51 - .75	5	3	6	5	5	1
Above .75	0	0	0	0	0	0

*Three of these values were exactly .25

Summary of Percentages of Students Answering Items Correctly

Range	Numbers of Items with Values in Each Range								
	Grade 5			Grade 8			Grade 11		
	Tryout Sample	Validation Sample	Statewide Sample	Tryout Sample	Validation Sample	Statewide Sample	Tryout Sample	Validation Sample	Statewide Sample
.25 or below	0	0	0	1	2	0	1	0	0
.26 - .50	15	3	12	14	13	22	9	6	8
.51 - .75	16	17	21	15	14	9	18	18	23
Above .75	2	13	0	2	3	1	4	8	1
Mean p Value	.54	.70	.56	.52	.52	.47	.59	.65	.56

RATING OF CRITICAL THINKING OF 11TH GRADERS *

We would like you to rate the critical thinking abilities of some of your 11th grade students.

For the purpose of these ratings, consider critical thinking to be the ability to deal with problems in a logical, systematic way. Some of the components of critical thinking are:

1. The tendency to be rational (is adept at being analytical; attacks problems in a logical manner).
2. Ability to draw proper inferences from a written passage, chart or table.
3. Ability to seek out and find relevant and reliable information needed to solve a problem.
4. Ability to draw conclusions (capable of organizing and interpreting facts in order to arrive at acceptable conclusions).

In the spaces below list 5 to 10 eleventh grade students who you feel have a high level of critical thinking ability. Similarly list 5 to 10 eleventh grade students whose critical thinking ability is at a low level. If you are unable to identify 5 students in each category, you may list fewer than 5.

Beside the name of each student you list, indicate what subjects you teach or have taught the student, if any. Also, identify the program the student is enrolled in if you know this.

	Student Name	Subject You Teach Student	Student's Program, e.g., Academic, General, Vocational
HIGH GROUP	1.		
	2.		
	3.		
	4.		
	5.		
	6.		
	7.		
	8.		
	9.		
	10.		
LOW GROUP	1.		
	2.		
	3.		
	4.		
	5.		
	6.		
	7.		
	8.		
	9.		
	10.		

*A similar rating instrument was used at the grade 8 level and in District 1 at the grade 5 level.

Teacher's Name _____ School _____

RATING OF CRITICAL THINKING (Grade 5)*

We would like you to rate the critical thinking abilities of your students.

For the purpose of these ratings, consider critical thinking to be the ability to deal with problems in a logical, systematic way. Some of the components of critical thinking are:

1. The tendency to be rational (is adept at being analytical; attacks problems in a logical manner).
2. Ability to draw proper inferences from a written passage, chart or table.
3. Ability to seek out and find relevant and reliable information needed to solve a problem.
4. Ability to draw conclusions (capable of organizing and interpreting facts in order to arrive at acceptable conclusions).

Below please list your grade 5 students. Then, rate each student by placing a check (✓) under the heading which best describes that student's critical thinking ability.

	Critical Thinking Ability				
Student Name	Outstandingly High	Somewhat Above Average	Average	Somewhat Below Average	Very Much Below Average

*This instrument was used only in Districts 3 and 4.